

Interoperability Session

BRC Kick-off Meeting
Oct 13-14 2004

Bruno Sobral (sobral@vbi.vt.edu)
Virginia Bioinformatics Institute
<http://www.vbi.vt.edu/>



Agenda

- **Presentation by Valentina Di Francesco (NIAID)**
- Presentation by me
- Round Table Discussion on Topics Sent Out
- Break
- Round Table Discussion: Establishing of BRC IOWG

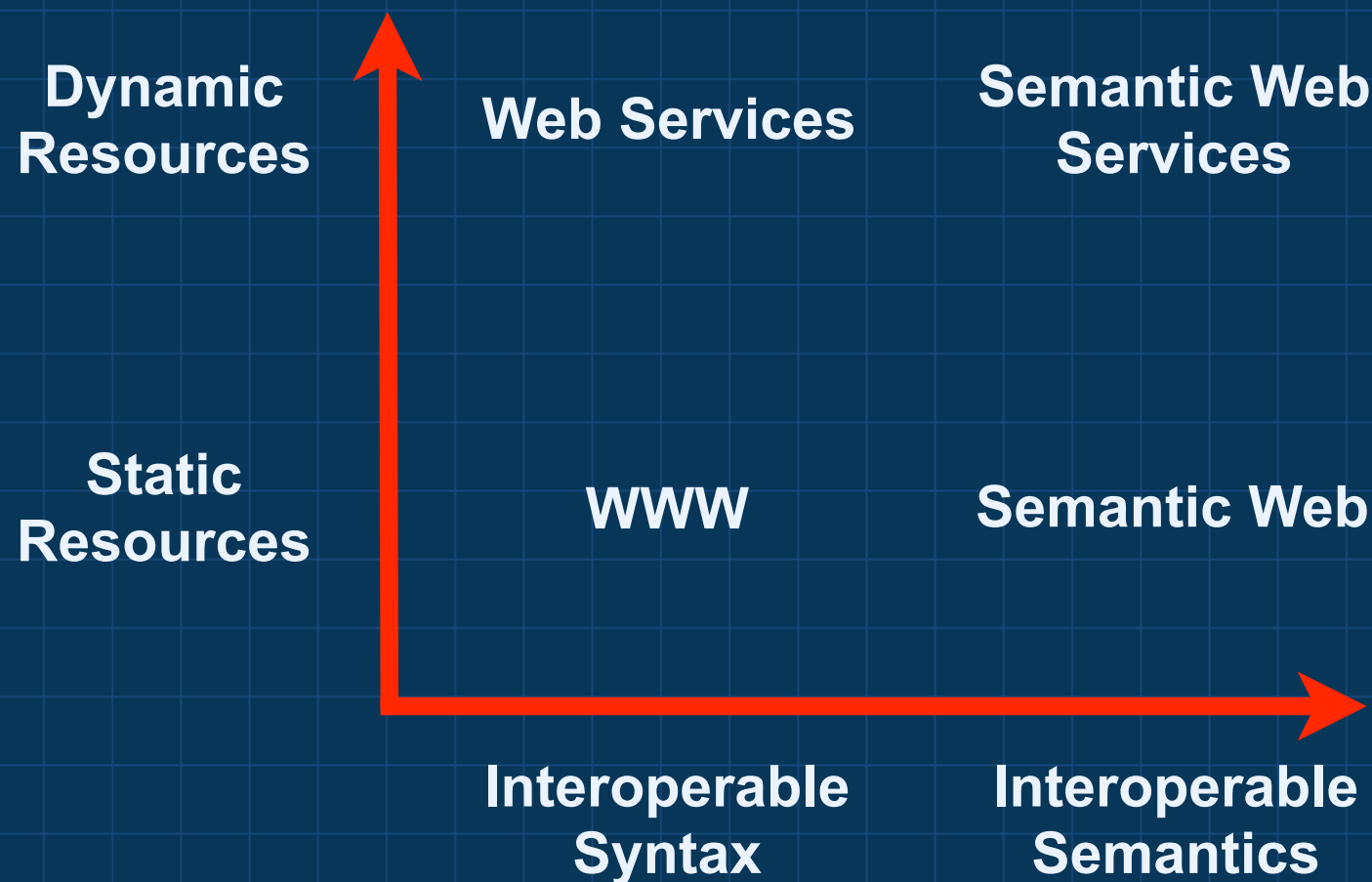
Agenda

- Presentation by Valentina Di Francesco (NIAID)
- **Presentation by me**
- Round Table Discussion on Topics Sent
- Break
- Round Table Discussion: Establishment of BRC's IOWG

What Is Meant By Interoperability?

- “...suitable for and capable of being implemented in a neutral manner on multiple operating systems and in multiple programming languages.” (WS-I)
- Deals with syntax, *not* semantics (see Host-Pathogen Interaction Session, under Ontology)
 - Syntax *required* to achieve integration (needs semantics as well)
- Focus: establishing machine-readable interfaces for key BRC components

Semantic Web Services

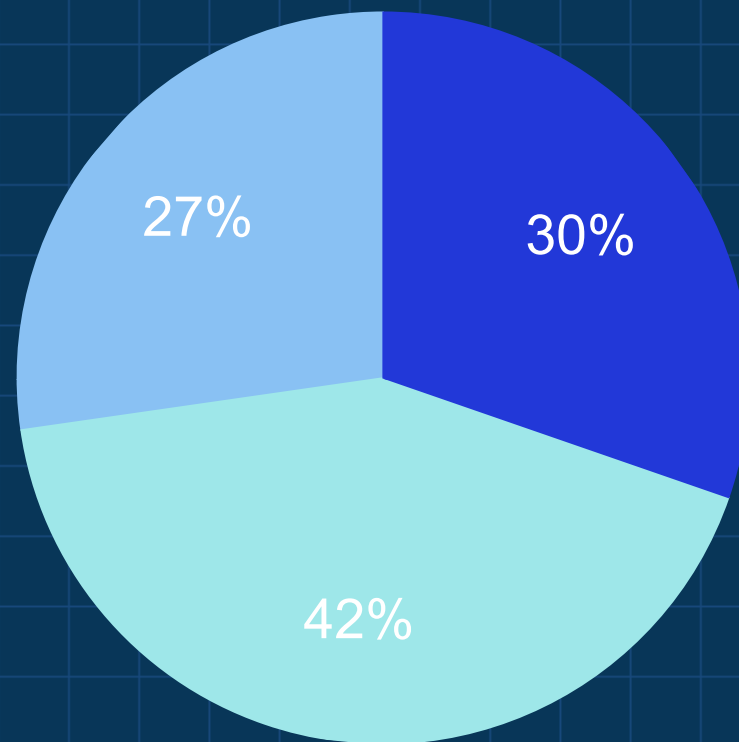


Why Does Interoperability Matter to BRCs?

- Imagine needing to do [comparative genomics] across bacterial (or viral) species whose data are contained in multiple BRCs....
- Imagine trying to integrate great (software) components from other BRCs....
- Imagine trying to integrate BRC systems if they are built as “proprietary” individual systems....
- Imagine trying to analyze host responses to diverse pathogens in which data are distributed across BRCs (and RCEs)....
- Etc....

Lack of Interoperability Is Costly: Average Cost/Business in 2001 = \$3.5 Million

● Software Costs ● Internal IT Staff ● Consulting Services



Source: Forrester Research

Cyberinfrastructure Characteristics



Cyberinfrastructure Characteristics

- Community-Focused
 - virtual organizations
 - distributed,
 - collaborative
- Scale and Scope
 - Multidisciplinary
 - International
 - Supporting data- and compute-intensive applications
 - High-end to desktop
 - Heterogeneous
- Common Technology & Policy Platform(s)
 - Interoperability
 - Supports characteristics above



Debbie Crawford, NSF

(Some of) The Challenges

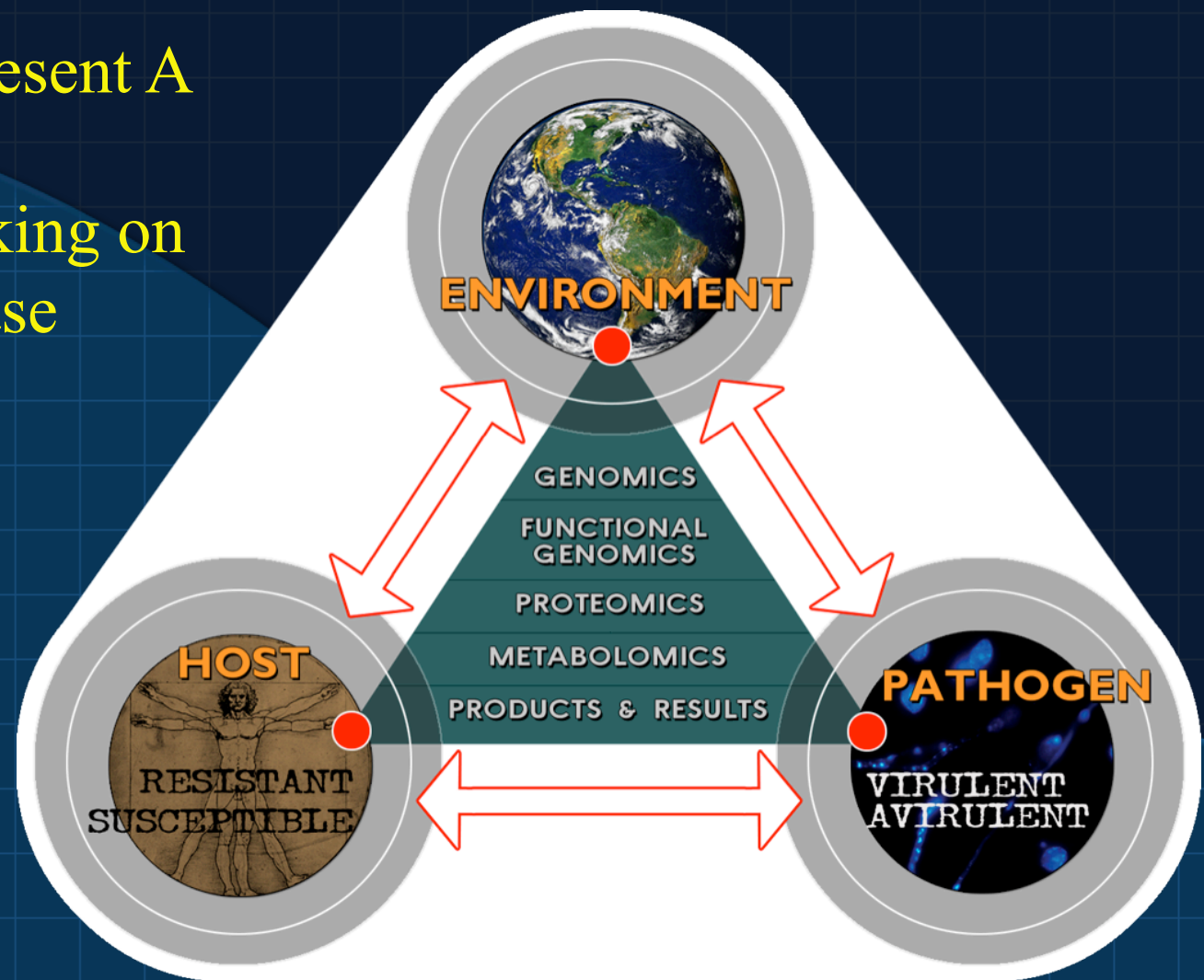


Challenging Context

- Institutional & Infrastructural Ecology
 - Technological change more rapid than institutional change
- Broadening Participation
- Community-Building
- Seamless Integration of New and Old
 - Balancing upgrades of existing and creation of new resources
 - Legacy data/models
- Providing sustainable support

PathoSystems Biology

The BRCs Represent A
Component of a
Federation Working on
Infectious Disease
Research



Role of IT in PathoSystems Biology

BIOTERRORISM

Information Technology Strategy Could Strengthen Federal Agencies' Abilities to Respond to Public Health Emergencies

What GAO Found

The six key federal agencies involved in bioterrorism preparedness and response identified about 70 planned and operational information systems in several IT categories associated with supporting a public health emergency. These encompass detection (systems that collect and identify potential biological agents from environmental samples), surveillance (systems that

facilitate
related
delivery
support
other c



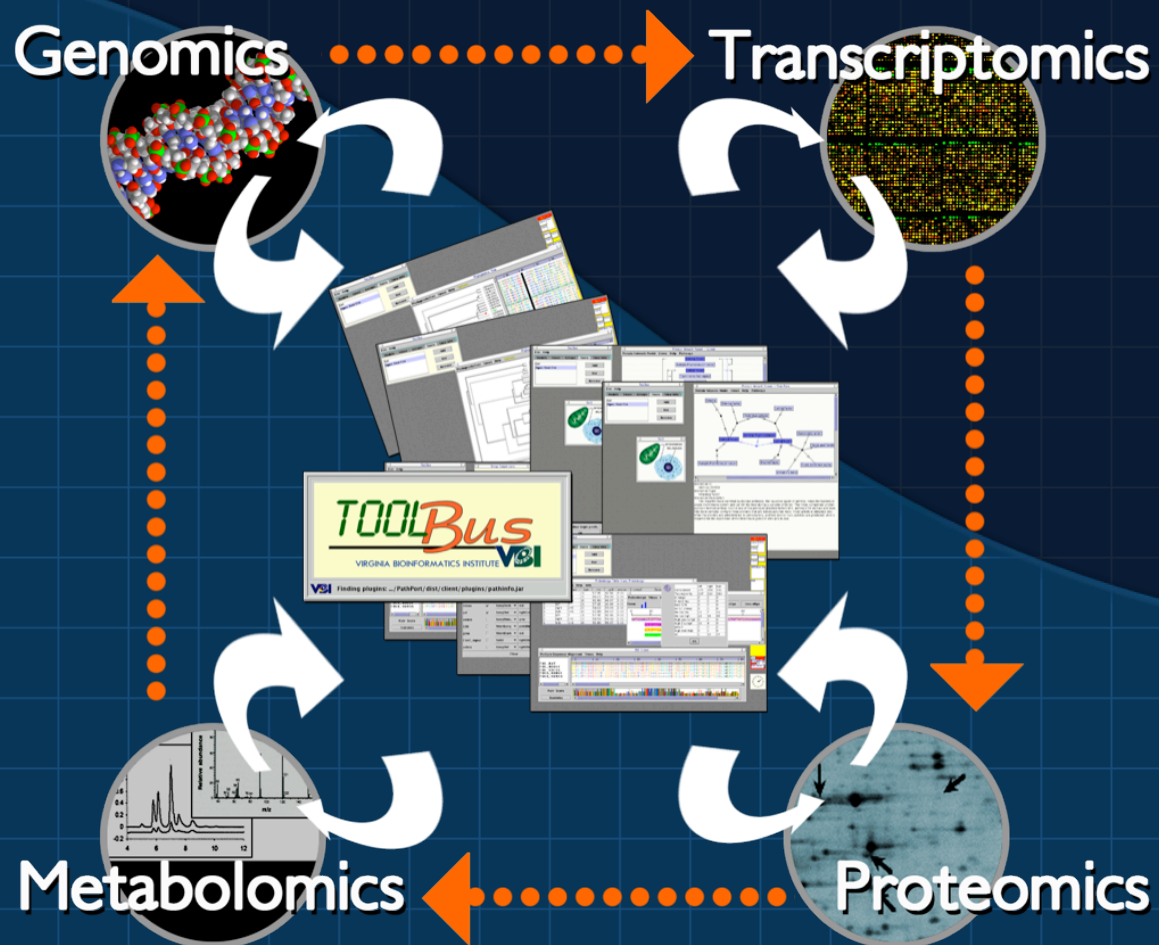
GAO Report,
May 2003, 03-139

Summary of the Systems Inventory by Agency

IT Categories	HHS	Defense	Energy	Agriculture	EPA	VA	Total
Detection	0	4	6	0	0	0	10
Surveillance	18	7	2	6	0	1	34
Communications	5	2	0	3	0	0	10
Supporting Tech	5	1	6	1	5	0	18
Total	28	14	14	10	5	1	72

Source: GAO.

Systems Interoperation: An Anecdote from the Pathogen Portal Project



Build on Open Standards

- Common vocabulary: Gene Ontology (GO)
- Transport format: XML
- Data definition language: XSD
- Wire protocol: SOAP
- Service definition language: WSDL
- Service registry: UDDI

After L. Stein, CSHL

PathPort's Use of Community XMLs

- Utilizes established, open community standards
 - *DAS-ML, BSML, MSA-ML (DNA) - Year 1*
 - *MAGE-ML (mRNA profiling) - Year 2*
 - *PEDRo (protein profiling) - Year 3*
 - *SBML (molecular models) - Year 3*
 - CellML (cellular levels including metabolism and signal transduction) - Year 4
 - AnatML (organ levels) - Year 4
 - FieldML (spatially and temporally varying field information using finite elements) - Year 5

ToolBus

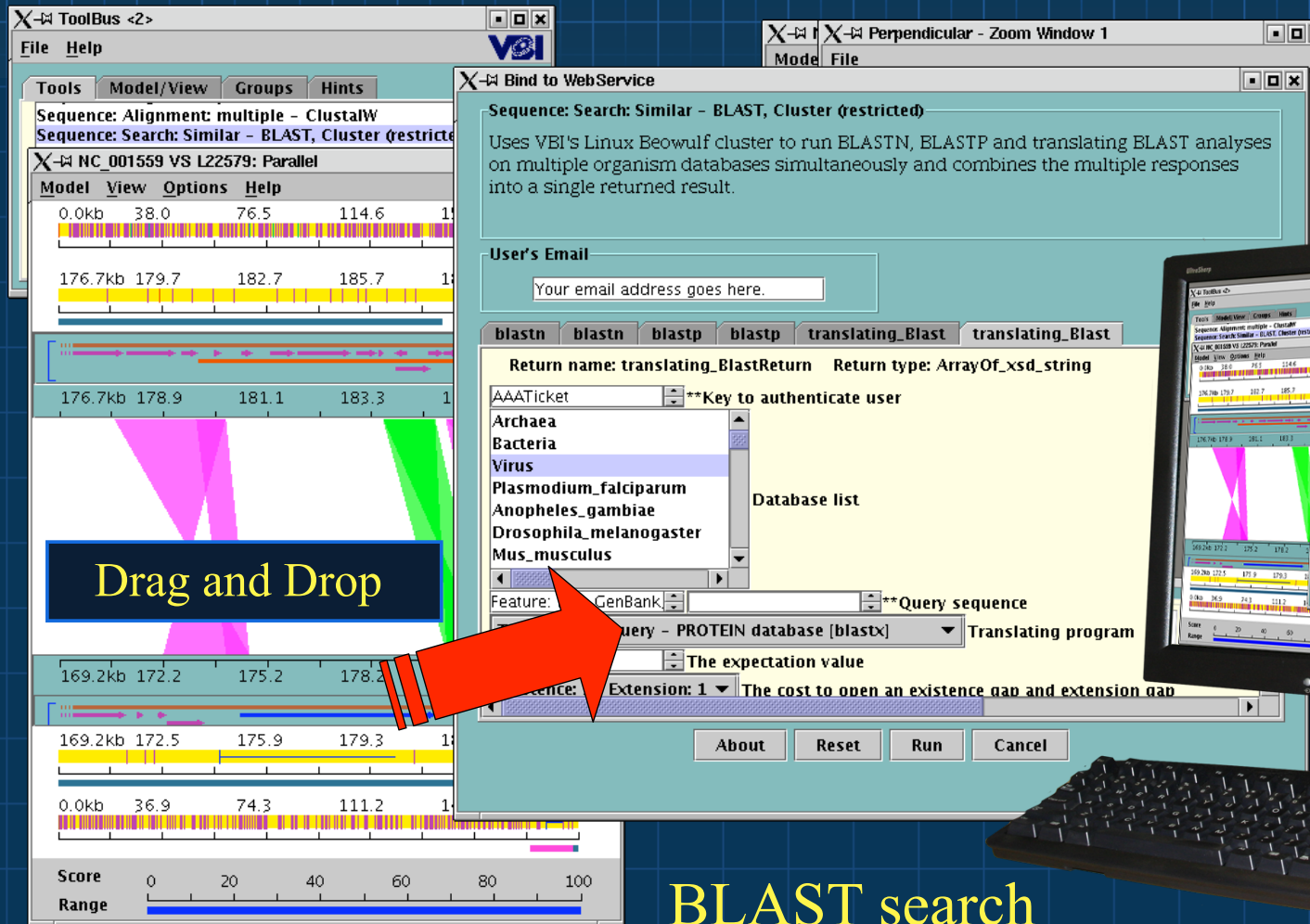
A client-side interconnect with the following goals:

- Platform independent
- Easily extensible
- Allow user-defined associations
- Easy to use



PathPort Comparative Genomics Scenario

step 07



ToolBus <2>

File Help

Tools Model/View Groups Hints

Sequence: Alignment: multiple - ClustalW
Sequence: Search: Similar - BLAST, Cluster (restricted)

NC_001559 VS L22579: Parallel

Model View Options Help

0.0kb 38.0 76.5 114.6 118.0

176.7kb 179.7 182.7 185.7 188.0

176.7kb 178.9 181.1 183.3 185.0

169.2kb 172.2 175.2 178.2 181.2

169.2kb 172.5 175.9 179.3 182.3

0.0kb 36.9 74.3 111.2 148.0

Score Range 0 20 40 60 80 100

Bind to Webservice

Sequence: Search: Similar - BLAST, Cluster (restricted)

Uses VBI's Linux Beowulf cluster to run BLASTN, BLASTP and translating BLAST analyses on multiple organism databases simultaneously and combines the multiple responses into a single returned result.

User's Email

Your email address goes here.

blastn blastn blastp blastp translating_Blast translating_Blast

Return name: translating_BlastReturn Return type: ArrayOf_xsd_string

AAATicket **Key to authenticate user

Archaea
Bacteria
Virus
Plasmodium_falciparum
Anopheles_gambiae
Drosophila_melanogaster
Mus_musculus

Database list

Feature: GenBank **Query sequence

Query - PROTEIN database [blastx] Translating program

The expectation value

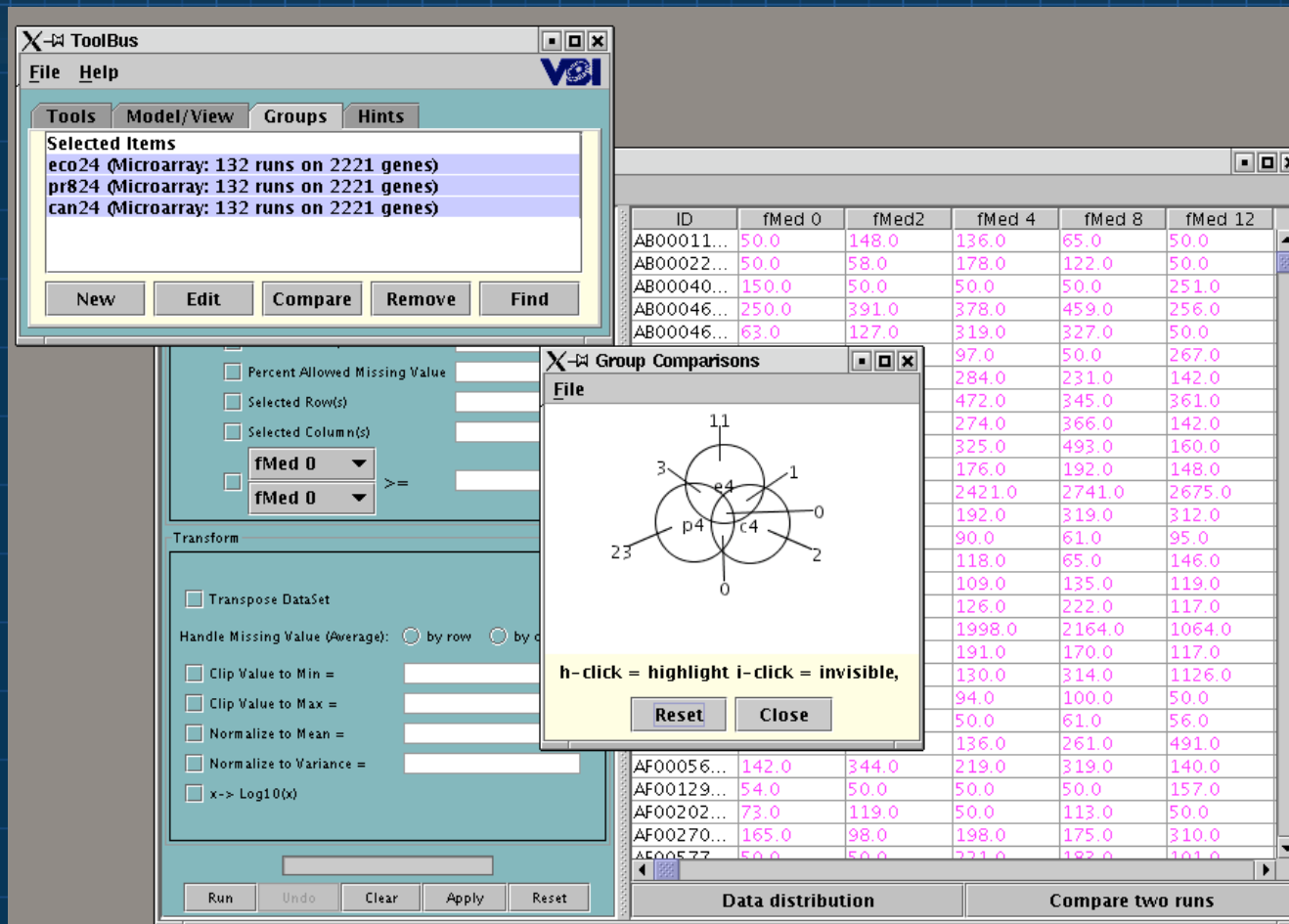
Extension: 1 The cost to open an existence gap and extension gap

About Reset Run Cancel

Drag and Drop

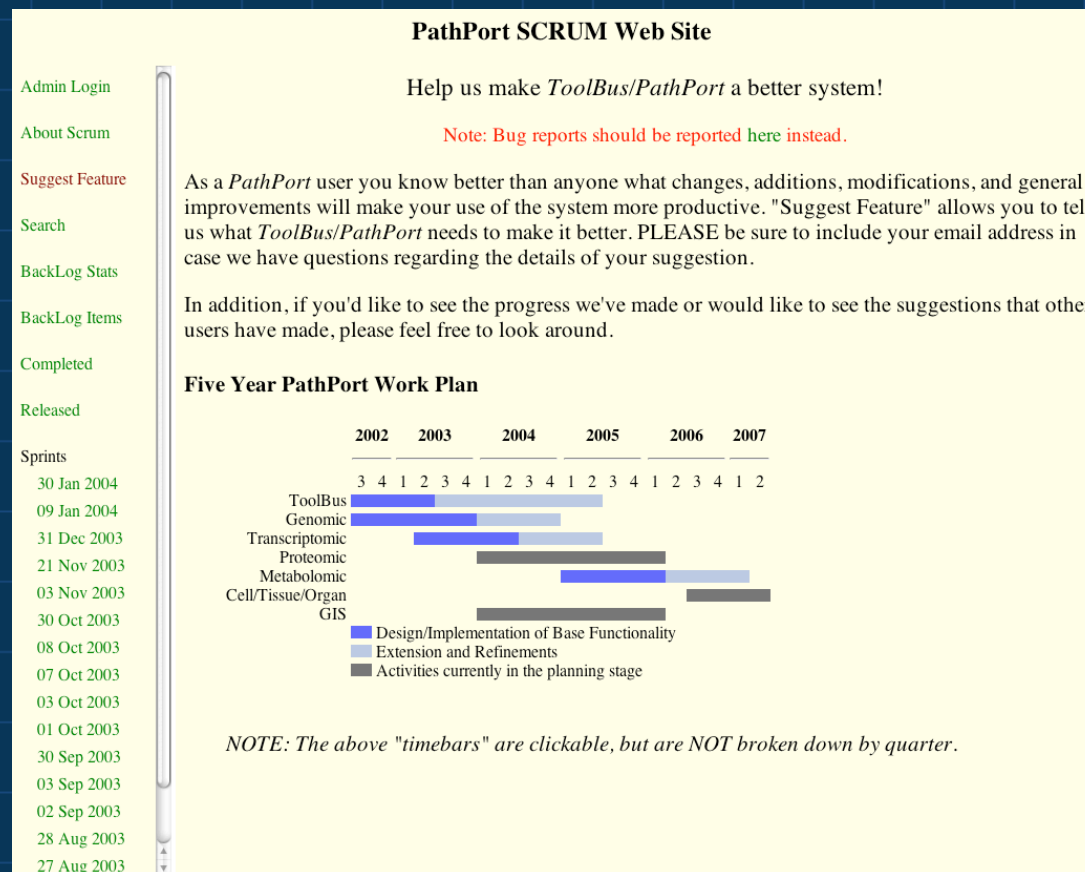
BLAST search

PathPort Group Comparison Scenario Seeking Diagnostic Markers



Comparison of three groups

SCRUM: Focus on Product, not Process



SCRUM Provides Interface for Community Requests

PathPort SCRUM Web Site

[Admin Login](#)

[About Scrum](#)

[Suggest Feature](#)

[Search](#)

[BackLog Stats](#)

[BackLog Items](#)

[Completed](#)

[Released](#)

[Sprints](#)

[30 Jan 2004](#)

[09 Jan 2004](#)

[31 Dec 2003](#)

[21 Nov 2003](#)

[03 Nov 2003](#)

Suggest Feature

Category:

Sub-Category: (required if chosen category has any subcategories)

Feature Name: (e.g., feature1)

Requestor's Name: (e.g., Dana Eckart)

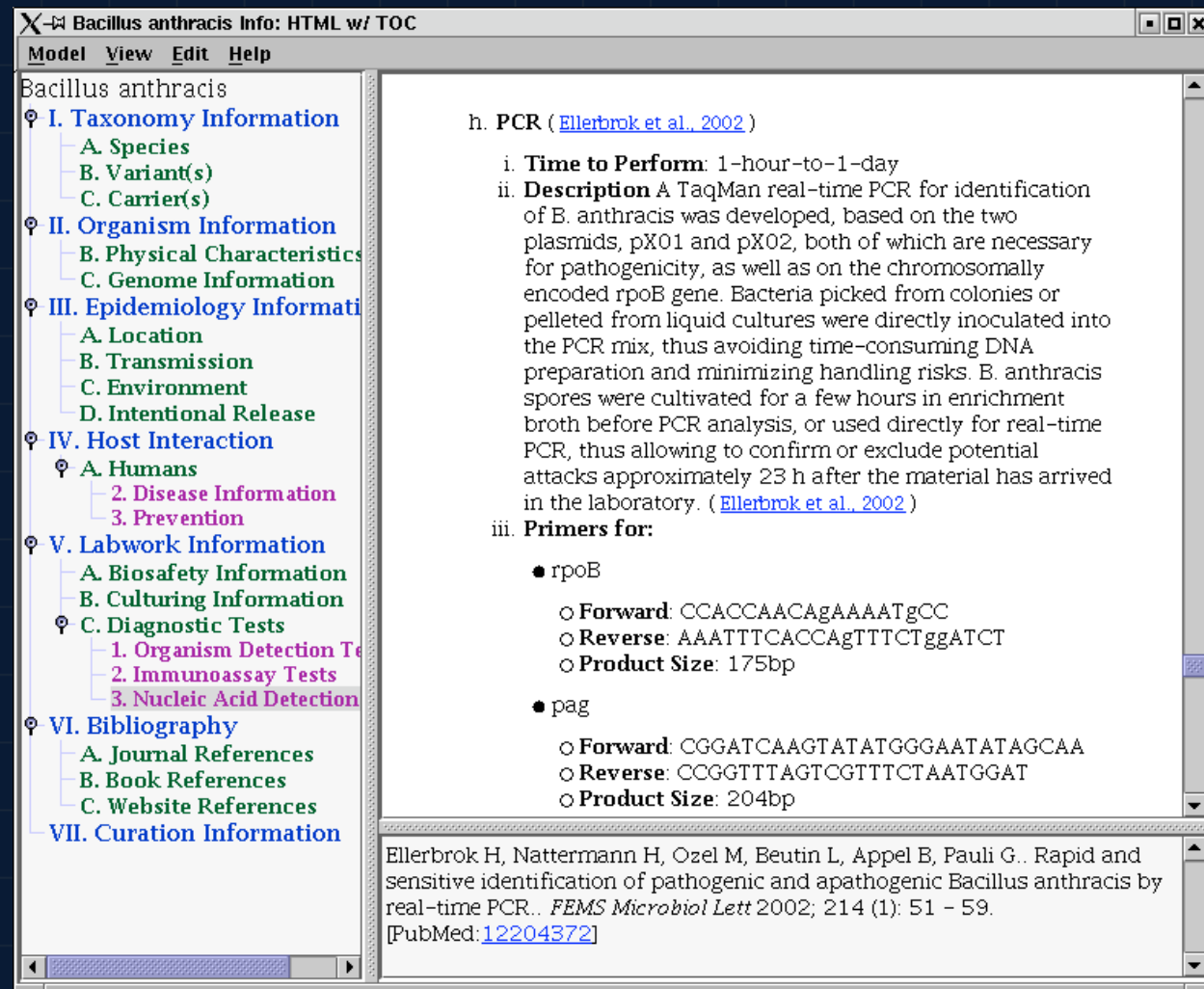
E-mail: (e.g., dana@vbi.vt.edu)

Providing your email address allows us to update you as to the status of your suggestion and indicates your willingness to receive such reports.

Detailed Description: *

Literature Curation: Pathogen Background Information

PathInfo Viewer



The screenshot shows a web browser window titled "Bacillus anthracis Info: HTML w/ TOC". The browser has a menu bar with "Model", "View", "Edit", and "Help". On the left side, there is a table of contents (TOC) for "Bacillus anthracis" with the following sections:

- I. Taxonomy Information
 - A. Species
 - B. Variant(s)
 - C. Carrier(s)
- II. Organism Information
 - B. Physical Characteristics
 - C. Genome Information
- III. Epidemiology Information
 - A. Location
 - B. Transmission
 - C. Environment
 - D. Intentional Release
- IV. Host Interaction
 - A. Humans
 - 2. Disease Information
 - 3. Prevention
- V. Labwork Information
 - A. Biosafety Information
 - B. Culturing Information
 - C. Diagnostic Tests
 - 1. Organism Detection Tests
 - 2. Immunoassay Tests
 - 3. Nucleic Acid Detection
- VI. Bibliography
 - A. Journal References
 - B. Book References
 - C. Website References
- VII. Curation Information

The main content area on the right displays the following information:

- h. PCR ([Ellerbrok et al., 2002](#))
 - i. **Time to Perform:** 1-hour-to-1-day
 - ii. **Description** A TaqMan real-time PCR for identification of *B. anthracis* was developed, based on the two plasmids, pX01 and pX02, both of which are necessary for pathogenicity, as well as on the chromosomally encoded *rpoB* gene. Bacteria picked from colonies or pelleted from liquid cultures were directly inoculated into the PCR mix, thus avoiding time-consuming DNA preparation and minimizing handling risks. *B. anthracis* spores were cultivated for a few hours in enrichment broth before PCR analysis, or used directly for real-time PCR, thus allowing to confirm or exclude potential attacks approximately 23 h after the material has arrived in the laboratory. ([Ellerbrok et al., 2002](#))
 - iii. **Primers for:**
 - *rpoB*
 - **Forward:** CCACCAACAgAAAATgCC
 - **Reverse:** AAATTTCACCAgTTTCTggATCT
 - **Product Size:** 175bp
 - *pag*
 - **Forward:** CGGATCAAGTATATGGGAATATAGCAA
 - **Reverse:** CCGGTTTAGTCGTTTCTAATGGAT
 - **Product Size:** 204bp

At the bottom of the main content area, the following text is displayed:

Ellerbrok H, Nattermann H, Ozel M, Beutin L, Appel B, Pauli G.. Rapid and sensitive identification of pathogenic and apathogenic *Bacillus anthracis* by real-time PCR.. *FEMS Microbiol Lett* 2002; 214 (1): 51 - 59.
[PubMed: [12204372](#)]

Pathogen Information Markup Language (PIML)

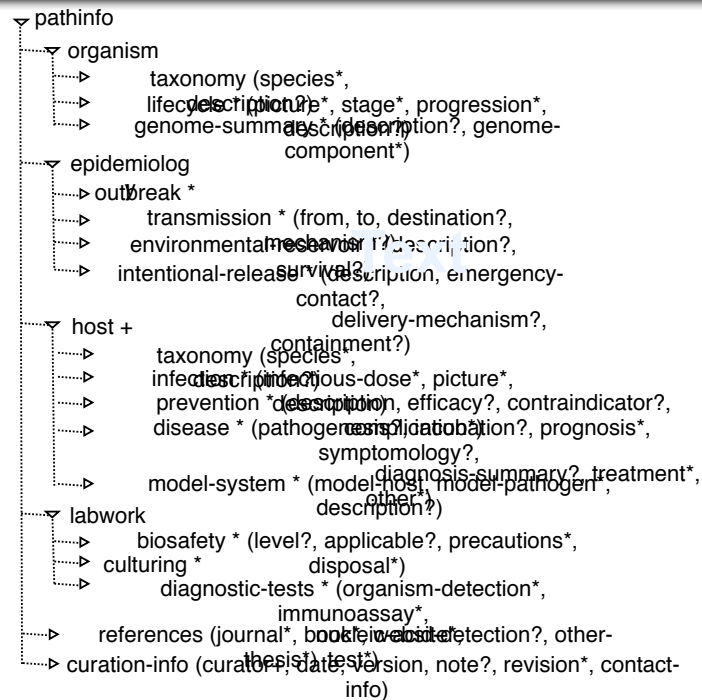
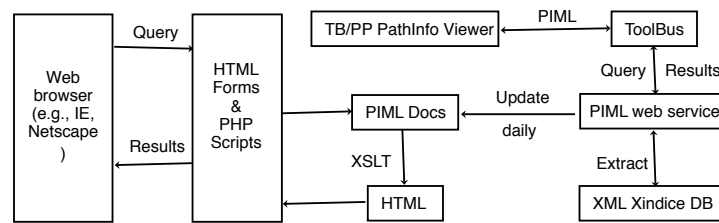


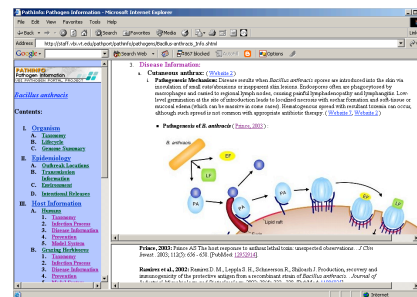
Figure 1. The PIML structure overview. The lower levels of elements and all element attributes are not shown but can be viewed from the PathInfo project website. A particular element can appear one or more times, zero or more times, or zero or one times, signaled by the plus sign (+), asterisk (*), and question mark (?), respectively.

PathInfo Query and Display System

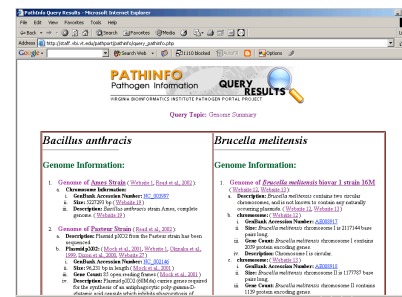
THE PATHOGEN PORTAL
https://www.vbi.vt.edu/pathport



(A)
Architecture



(B) Complete Pathogen
Information



(C) Comparison of Pathogens

He et al., Bioinformatics, in press

Agenda

- Presentation by Valentina Di Francesco (NIAID)
- Presentation by me
- **Round Table Discussion on Topics Sent**
- Break
- Round Table Discussion: Establishment of BRC's IOWG

Agenda

- Presentation by Valentina Di Francesco (NIAID)
- Presentation by me
- Round Table Discussion on Topics Sent
- **Break**
- Round Table Discussion: Establishment of BRC's IOWG

Agenda

- Presentation by Valentina Di Francesco (NIAID)
- Presentation by me
- Round Table Discussion on Topics Sent
- Break
- **Round Table Discussion: Establishment of BRC's IOWG**